

Inquiry of Personal Intelligence Of Adolescence and Early Adulthood College Students of Tamilnadu using Id3 Algorithm

S.Muthukumar, P.Geetha, E.Ramaraj

ABSTRACT: Educational Data Mining is an Emerging research domain which focus on extracting knowledge from educational databases to promote the learning environment. Educational psychology helps to understand the differences in learning process of a person from cognitive and behavioral perspective. This research aims to predict the personal intelligence of adolescence and early adulthood who are studying under graduate and post graduate courses in educational institution. Adolescence plays a very important role in the development of personality to new dimensions in human life. The Adolescence physical, mental, social, moral and spiritual outlooks undergo revolutionary changes. Many teachers and parents fail to assess these changes and they do not like to slacken their control over them. Psychologist has stressed to properly channelize the behavior of adolescence and give them adequate education. In this paper machine learning technique with decision tree induction algorithm was used to analyze the personal intelligence of college students. To construct the decision tree, Entropy and Information Gain are used as attribute selection measures in ID3 algorithm. Applying data mining techniques in the field of Educational Psychology is a new method and the rules generated from the decision tree helps to identify the personal intelligence of the students. This result helps the Educators to improve the learning environment better for Adolescence.

Keywords: Educational Psychology, Educational Data Mining, Decision Tree Induction Algorithm.

I. INTRODUCTION

Psychology is the scientific study of behavior and mental process and it is applied in many fields including Education. Educational psychology is the branch of Applied Psychology which applies the psychological rules, principles, theories and techniques to teaching-learning process. It has developed the procedures to identify the issues, collect the data, process the data and predict the behavior in a scientific way. The focusing area of Educational Psychology is the learner, learning process, learning experience, learning environment, teacher and teaching. Intelligence is the ability of a person to direct his behavior towards a goal[1]. Personal intelligence refers to the mental ability of a person to think and act towards the efficient adjustment in the environment. The personal intelligence of a person emerges from the brain and act according to the situation present in the outside world.

Revised Manuscript Received on August 05, 2019.

S.Muthukumar, Research Scholar, Department of Computer Science, Alagappa University, Karaikudi, India.

P.Geetha, Associate Professor, Department of Computer Science, Dr. Umayal Ramanathan College for Women, Karaikudi, India.

E.Ramaraj, Professor and Head, Department of Computer Science, Alagappa University, Karaikudi, India.

The personal Intelligence mainly depends on two factors called as Environment which covers all the outside factors that have acted on the individual since he began his life, whereas heredity covers all factors that are present within the individual since from the time of birth[2]. Applying data mining techniques in Education is called as Educational Data Mining. Various techniques used in data mining like classification, clustering and regression are used in the educational data and knowledge is extracted from the database. This extracted knowledge helps the educators to improve the learning environment.

II. RELATED WORK

A.Talaie[3] proposed a new method to identify the icons used to represent the polymers used in chemical field. They constructed a decision tree with C4.5 algorithm using the current, mass and resistance parameters of the different chemicals and identified the polymers present in them. Catherine Butchart[4] used CART Tree to study the impatient mortality of oldest people have age greater than 90. Using the CART tree they identified the potential acute illness prognostic indicators responsible for inpatient mortality. Maricel Cabiling-Ramos[5] made a study to assess the attitude of preschoolers with respect to their gender, family nature, school nature, study habits and identified the behavior of the preschool children. K.Jamberi[6] proposed a method to prognosis Breast Cancer using the ID3 algorithm by Breast cancer image collected from medical centre. They combine Genetic algorithm with ID3 by using operators like selection, crossover and mutation in the breast cancer image and identified the image segments affected with breast cancer. L.R.Aravind Babu[7] had reviewed various machine learning approaches in Load Forecasting Methods. Several mathematical and computing techniques used in load forecasting are listed in this paper. Ravi Ranjan, D.Priyanka Sruti[8] in their paper used machine learning for choosing top managerial staff in a specific firm. They made calculations on firm's information, potential of chiefs, properties of chiefs etc to select the best person for the position of Board of Director for a particular company.

III. PROPOSED FRAMEWORK

The proposed architecture for the analysis of personal intelligence is shown in the Figure 1 which has five steps. First the objective of the problems faced by adolescence is discussed with psychologist, teachers, Doctors and Sports men and hence a questionnaire was framed.

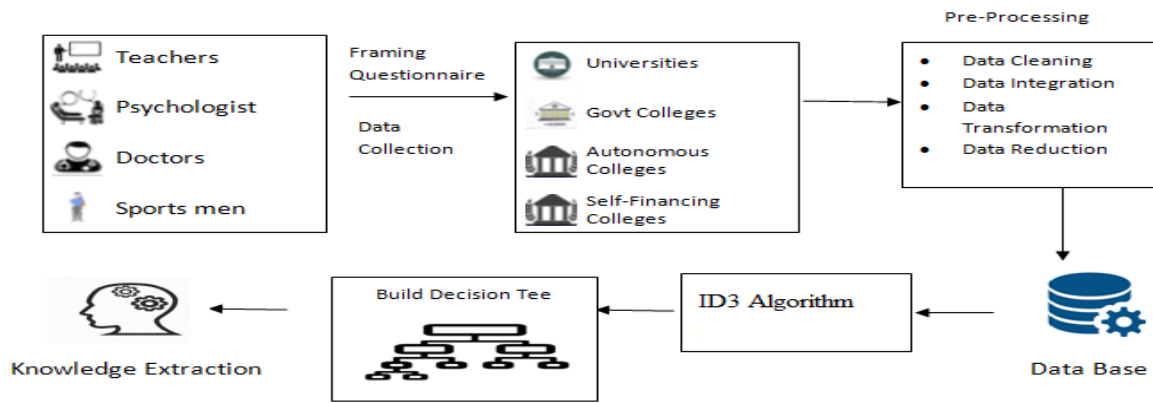


Figure 1: Framework of the Proposed Method

The questionnaire contains 27 questions based on the Eysenck’s Personality Inventory (EPI), and the problems discussed in various adolescence psychology books. The questionnaire is a closed type questionnaire having five point rating scale containing traits such as Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree. The second step of this research deals with data collected from various Universities, Government Colleges, Autonomous Colleges, Self-financing colleges. The questionnaire is given to 2083 Students who are Under Graduate and Post Graduate of Ramanathapuram, Sivagangai, Pudhukottai, Cuddalore, Villupuram and Kallakurichi Districts of Tamilnadu. From the 2083 students 720 were male and 1363 were female. The third step deals with pre-processing of data like data cleaning, data integration, data transformation, data reduction and finally the preprocessed data was stored in an Excel-2007. The fourth step of this research deals with implementing Machine Learning Techniques called Decision Tree Induction Algorithm in the Dataset The fifth step of this research deals with Rule Generation from the decision tree and extraction of the hidden knowledge from the dataset. The Extracted knowledge is given to the Educators to improve the Learning Environment of the Adolescence.

A. Feature Extraction for Decision Tree

A Feature is an attribute contained in a Dataset and the most influencing attribute from a dataset is chosen by a selection measure. Normally, Decision Tree has three Attribute Selection Measure namely Information Gain, Gain Ratio, Gini Index. For this research the Information Gain is used as the selection measure and the Pseudocode for constructing the decision tree is as shown below.

Input: Database D with details of student’s personal behavior

Output: A Decision tree containing the category of workers

Decision Tree_Creation(Dataset D, Target attribute C, Attributes A, Tuples T)

Begin

1. If D is Empty then
 2. return false
 3. Otherwise do
 4. Split the dataset into training and testing dataset
 5. If all the values in the target attribute in the training set are same then
 6. Return decision tree having single node with most common value of target attribute Ci
 7. Otherwise do
 8. **Foreach** possible value ai of A begin
 9. Compute Entropy for target attribute T using $Info(D)=\sum_{i=1}^m p_i \log_2 p_i$. Equ (1)
 10. Compute Information gain for each attribute using $InfoA(D)=-\sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j)$. Equ (2)
 11. Compute Gain for each attribute using $Gain(A)=Info(D)-Info_A(D)$ (3) Equ (3)
 - end
 12. Select attribute with maximum Information gain as root node and split the values as individual branches
 13. Repeat the steps until the desired tree was generated
- End

IV. RESULTS OBTAINED FOR STUDENT PERSONAL INTELLIGENCE DATASET

For the implementation of Decision tree Algorithm the open source version of Rapid Miner9.1 was used. The parameters used to build the decision tree in Rapid Miner are as follows. The Criterion is Information gain, minimum size for split is 100, and minimal leaf size is 10, minimal gain is 0.5.

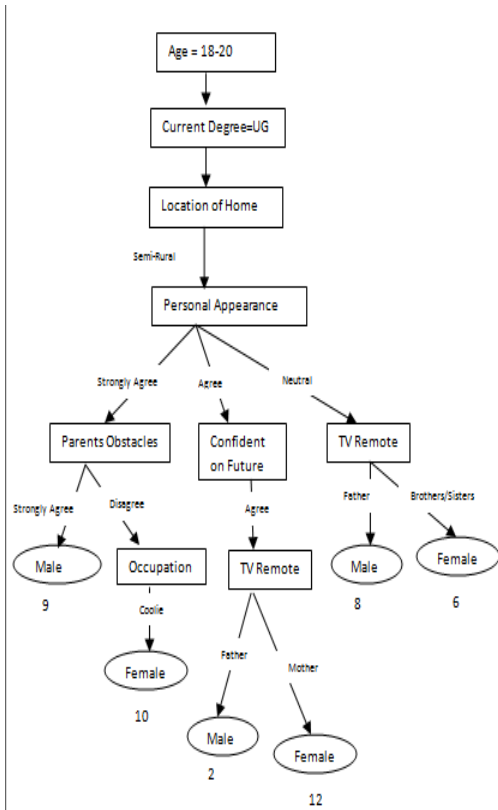


Figure 2(a): Decision Tree for Location Semi-Rural

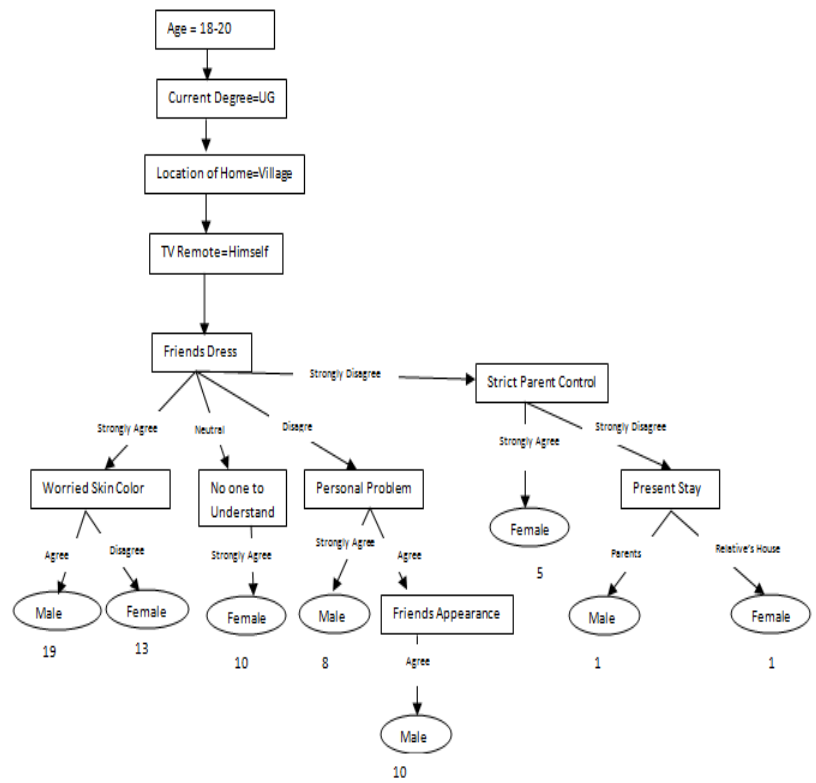


Figure 2(b): Decision Tree for Location Village

The Representation of decision tree for the student's personal intelligence dataset for students who are studied in Co-Education Schools and Under Graduate nature having Co-Education and Age belonging to 18-20 is shown in Figure 2(a) and Figure 2(b). The rules generated from Decision tree shown in Figure 2(a) are as follows: R1: If nature of School is Co-Education and UG nature is Co-Education and age is 18-20 and Current Degree is UG and Location of home is Semi-Rural and Personal Appearance is Strongly Agree and Parents as obstacles is Strongly Agree then Gender is male (9) tuples. R2: If nature of School is Co-Education and UG nature is Co-Education and age is 18-20 and Current Degree is UG and Location of home is Semi-Rural and Personal Appearance is Strongly Agree and Parents as obstacles is Disagree and Occupation is Coolie

then Gender is female (10) tuples. Like this 6 rules were generated from Figure 2(a). The rules generated from Decision tree shown in Figure 2(b) are as follows: R1: If nature of School is Co-Education and UG nature is Co-Education and age is 18-20 and Current Degree is UG and Location of home is Village and TV Remote is with himself and Worried about friends dress is Strongly Agree and Worried about skin color is Agree then Gender is male (19) tuples. R2: If nature of School is Co-Education and UG nature is Co-Education and age is 18-20 and Current Degree is UG and Location of home is Village and TV Remote is with himself and Worried about friends dress is Strongly Agree and Worried about skin color is Disagree then Gender is female (13) tuples. Like this 8 rules were generated from Figure 2(b).

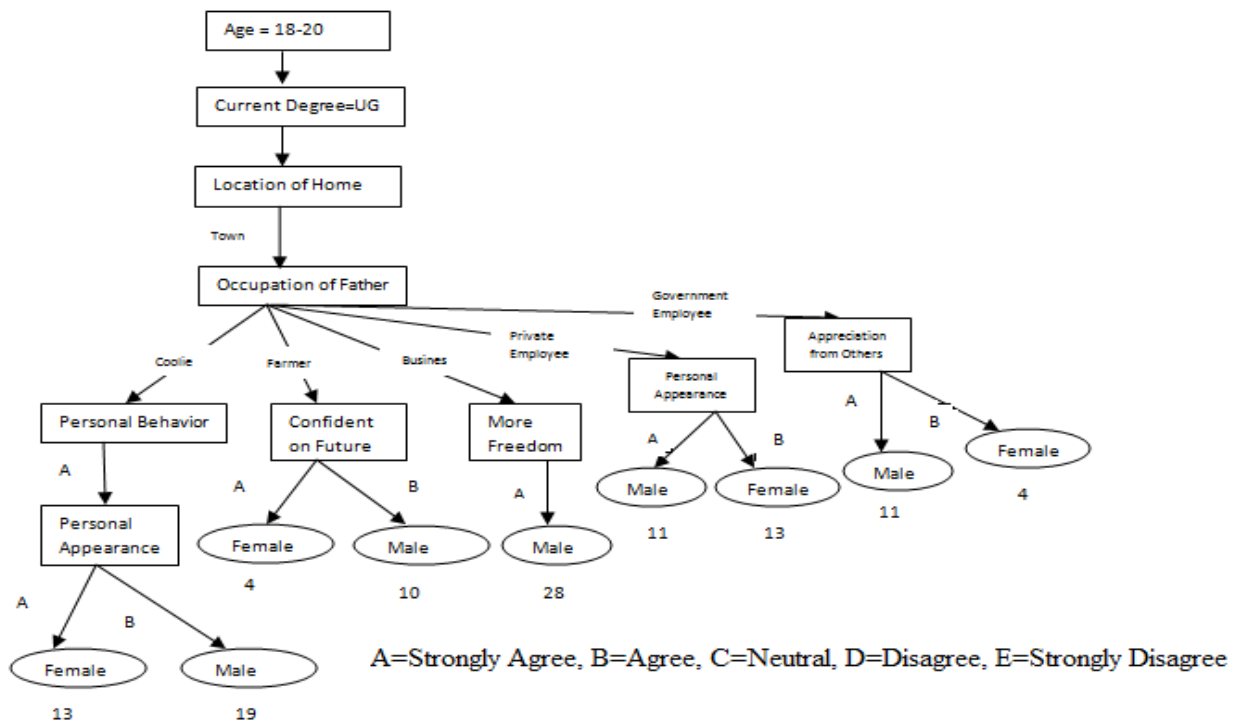


Figure 3: Decision Tree for Location of Home is Town

The rules generated from Decision tree shown in Figure 3 are as follows: R1: If nature of School is Co-Education and UG nature is Co-Education and age is 18-20 and Current Degree is UG and Location of home is Town and Occupation of Father is Coolie and Personal Behavior is Strongly Agree and Worried about Personal Appearance is Strongly Agree then Gender is female (13) tuples. Like these 9 rules were generated from Figure 3. The Decision Tree obtained for School nature Co-Education and Age 21-23 is shown in Figure 4. Rules generated from the above decision tree shown in Figure 4 are: R1:IF Nature of School is Co-Education and Currently Studying is UG and Worried about Skin Color is Strongly Agree then Gender=Female(5 tuples). R2:IF Nature of School is Co-Education and Currently Studying is UG and Worried about Skin Color is

Neutral and No one to Understand is Strongly Agree then Gender=Female(6 tuples). R3:IF Nature of School is Co-Education and Currently Studying is UG and Worried about Skin Color is Neutral and No one to Understand is Neutral then Gender=Male(6 tuples). R4:IF Nature of School is Co-Education and Currently Studying is UG and Worried about Skin Color is Disagree and More Freedom is Strongly Agree then Gender=Female(2 tuples). R5:IF Nature of School is Co-Education and Currently Studying is UG and Worried about Skin Color is Disagree and More Freedom is Agree then Gender=Male(9 tuples). R6:IF Nature of School is Co-Education and Currently Studying is UG and Worried about Skin Color is Strongly Disagree and Persona Behavior is Strongly Agree then Gender=Male(24 tuples).

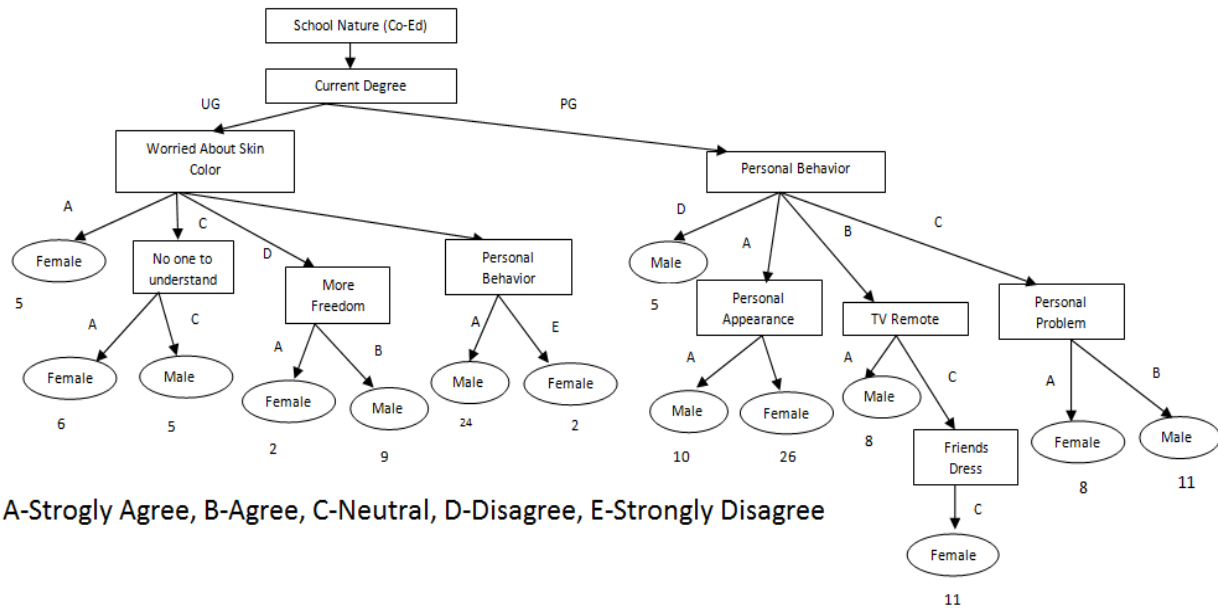


Figure 4: Decision Tee generated for Age [21-23]

R7:IF Nature of School is Co-Education and Currently Studying is UG and Worried about Skin Color is Strongly Disagree and Persona Behavior is Strongly Disagree then Gender=Female(2 tuples). R8:IF Nature of School is Co-Education and Currently Studying is PG and Personal Behavior is Disagree then Gender=Female(5 tuples). R9: IF Nature of School is Co-Education and Currently Studying is PG and Personal Appearance is Strongly Agree then Gender=Female (26 tuples). R10: IF Nature of School is Co-Education and Currently Studying is PG and Personal Appearance is Agree then Gender=Male (10 tuples). R11: IF Nature of School is Co-Education and Currently Studying is PG and TV Remote is with Father then Gender=Male (10 tuples). R12: IF Nature of School is Co-Education and Currently Studying is PG and TV Remote is with her then Worried about Friends dress is Neutral then Gender=Female (11 tuples). R13: IF Nature of School is Co-Education and Currently Studying is PG and Personal Problem is Strongly Agree then Gender=Female (8 tuples). R14: IF Nature of School is Co-Education and Currently Studying is PG and Personal Problem is Agree then Gender=Male (11 tuples).

$$= \left(\frac{720}{2083} \log_2 \frac{720}{2083} - \frac{1363}{2083} \log_2 \frac{1363}{2083} \right) = (0.3456 \times 1.5325 - 0.6543 \times 0.6116) = 0.9297$$

The information required to classify School nature attribute from the dataset on partitioning is calculated by $Info_A(D) = -\sum_{j=1}^V \frac{|D_j|}{|D|} \times Info(D_j)$

$$= \left(\frac{541}{2083} \right) \left(\frac{15}{541} \log_2 \frac{15}{541} - \frac{526}{541} \log_2 \frac{526}{541} \right) + \left(\frac{225}{2083} \right) \left(\frac{181}{225} \log_2 \frac{181}{225} - \frac{44}{225} \log_2 \frac{44}{225} \right) + \left(\frac{1317}{2083} \right) \left(\frac{524}{1317} \log_2 \frac{524}{1317} - \frac{793}{1317} \log_2 \frac{793}{1317} \right) = (0.0475 + 0.0231 + 0.0844) = 0.255$$

Gain for School Nature Attribute; Gain (A) = Info (D) – Info_A(D) = 0.9297 - 0.255 = 0.6747. Calculating the Information for the entire attribute present in the dataset by the same manner, comparing the Gain value with each attribute and the attribute with highest Information Gain is chosen as the root node for the Decision Tree.

A. EVALUATION OF RESULTS

The dataset is classified on the basis of Gender attribute, and the total number of records based on the gender attribute is given in Table1.

Table1: Details of Gender Attribute

Male	Female	Total
720	1363	2083

Table2: School Wise Details

School Nature	Male	Female	Total
Girls	15	526	541
Boys	181	44	225
Co-Education	524	793	1317
Total	720	1363	2083

B. PERFORMANCE MEASURES

The performance of a classifier used for classification is done by the confusion matrix obtained for the dataset. To Calculate the performance, the student dataset has been split into Training Dataset as 70% and Testing Dataset as 30%, and the following confusion matrix is obtained.

Table3: Performance of the ID3 Algorithm in Rapid Miner

	Female	Male	Class Precision
Female	330	73	81.89%
Male	79	143	64.41%
Class Recall	80.68%	66.20%	
Accuracy of the Classifier			75.68%

The information needed to classify based on Gender is calculated by $Info(D) = -\sum_{i=1}^m p_i \log_2 p_i$.



V. DISCUSSION

The Research clearly shows that personality of Adolescence mainly depends on the nature of the school. Students who had their Schoolings at Girls School and Undergraduate at Women's College are personally very Strong. The TV Remote attribute used in this Dataset is a symbolic representation of Control over the Family whether Father or Mother or Grandmother/Grandfather and it plays a very important role in the personality development of Girls. The personality of the boys who had their schoolings at Boys School and Undergraduate at Men's College is also personally very Strong and their personality depends on the occupation of the parents. Students who had their schoolings at Co-Education Schools and Currently Studying Under Graduate at Co-Education College having Age of 18-20 are personally very weak. The Age group 18-20 students have lot of personal problem that affects their personality. These Students care more for their personal appearance and worried more about the dress code. They fight for personal freedom with their parents and they think their parents as obstacles to fulfill their desires. The nature of the skin color and occupation of their parents also plays a significant role in shaping their personality. All these problems made them into depression. This research suggest that care must be taken by the parents and teachers of Age group 18-20 students and proper counseling must be given to the students to overcome these personal problem.

VI. CONCLUSION

This paper proposed a machine learning approach called ID3 Algorithm to predict the personal intelligence of adolescence in Tamilnadu and its performance are calculated using the performance metrics. The findings in this paper clearly show that the school and college environment plays a vital role in developing the personality of an individual. Students studying in Co-Education Schools and Colleges have lot of personal problems when compared with students studying in single gender Schools and Colleges. The students at the age group of 18-20 who are studying Co-Education have low self-esteem. It is suggested that all Educational Institutions must give proper counseling to students through psychologist to improve the self esteem of the students. Finally, the rules generated from the collected data helps to bridges the gap between the students with their staff and parents to improve the learning environment. In future, the proposed methodology will help to identify the emotional and social intelligence of adolescence.

ACKNOWLEDGMENTS

This article has been written with the financial support of RUSA-Phase 2.0 grant sanctioned vide Letter No.F.24-51/2014-U, Policy (TNMulti-Gen), Dep.of Edn.Govt. of India, Dt.09.10.2018.

REFERENCES

1. S.P. Chaube, "Psychology of Adolescents in India", Concept Publishing Company, New Delhi, 2002.
2. Dr.M.Manivannan, "Psychology of Learning and Human Development", Neelkamal Publishing, 2012.
3. Talaie, A., et al. "Application of an engineering algorithm with software code (C4. 5) for specific ion detection in the chemical

- industries." Advances in Engineering Software, Volume -30, Issue-1, pp:13-19, 1999.
4. Butchart, Catherine, et al. "Identification of possible determinants of inpatient mortality using Classification and Regression Tree (CART) analysis in hospitalized oldest old patients." Archives of gerontology and geriatrics, Volume-56, Issue-1, pp: 188-191, 2013.
5. Maricel Cabiling-Ramos, "Assessment on the Behavior and Attitude of the Preschoolers of La Consolacion University Philippines as Perceived by their Teachers", International Journal of Scientific & Engineering Research, Volume 9, Issue 1, January-2018.
6. K.Jamberi, E.Ramaraj, "Multilevel Classification Algorithm using Diagnosis and Prognosis of Breast Cancer", International Journal of Data Mining Techniques and Applications, Volume: 06, Issue: 02, December 2017, Page No.56-58.
7. L.R.Aravind Babu, "Effect of Artificial Neural Network Approach in Load Forecasting Methods", International Journal of Innovative Technology and Exploring Engineering ,Volume-8, Issue-8s, pp:38-41, 2019
8. Ravi Ranjan,D. Priyanka Sruti,S. Delfin, Aditya Kesharwa, "Selecting Board Directors using Machine Learning", International Journal of Innovative Technology and Exploring Engineering ,Volume-8, Issue-7C2, pp:39-41, 2019.

AUTHOR'S PROFILE



S.Muthukumaran is a Research Scholar in the Department of Computer Science, Alagappa University, Karaikudi. His main Research Interest includes Data Mining, Big Data Analytics, and Internet of Things. He has published 9 International Journals.



P.Geetha is working as an Associate Professor in the Department of Computer Science at Dr. Umayal Ramanathan College for women, Karaikudi. Her Research Interest includes Data Mining, Network Security and Internet of Things. She has published 14 International Journals.



E. Ramaraj is working as the Professor and Head of the Department of Computer Science, Alagappa University, Karaikudi. He has the sound knowledge in many research fields especially in Data Mining, Network Security, Remote Sensing, and Big Data & Analytics. He has published more than 100 international journals.